



CENTRO**ALGORITMI**

Interoperabilidade global dos Repositórios Digitais -
desafios à comunidade Ibero-Americana de metadados

Ana Alice Baptista



CENTROALGORITMI

Perspetiva





CENTROALGORITMI

O que me motiva?

(this is a political statement)

The Open Definition: “Open data and content can be freely used, modified, and shared by anyone for any purpose”



OPEN KNOWLEDGE

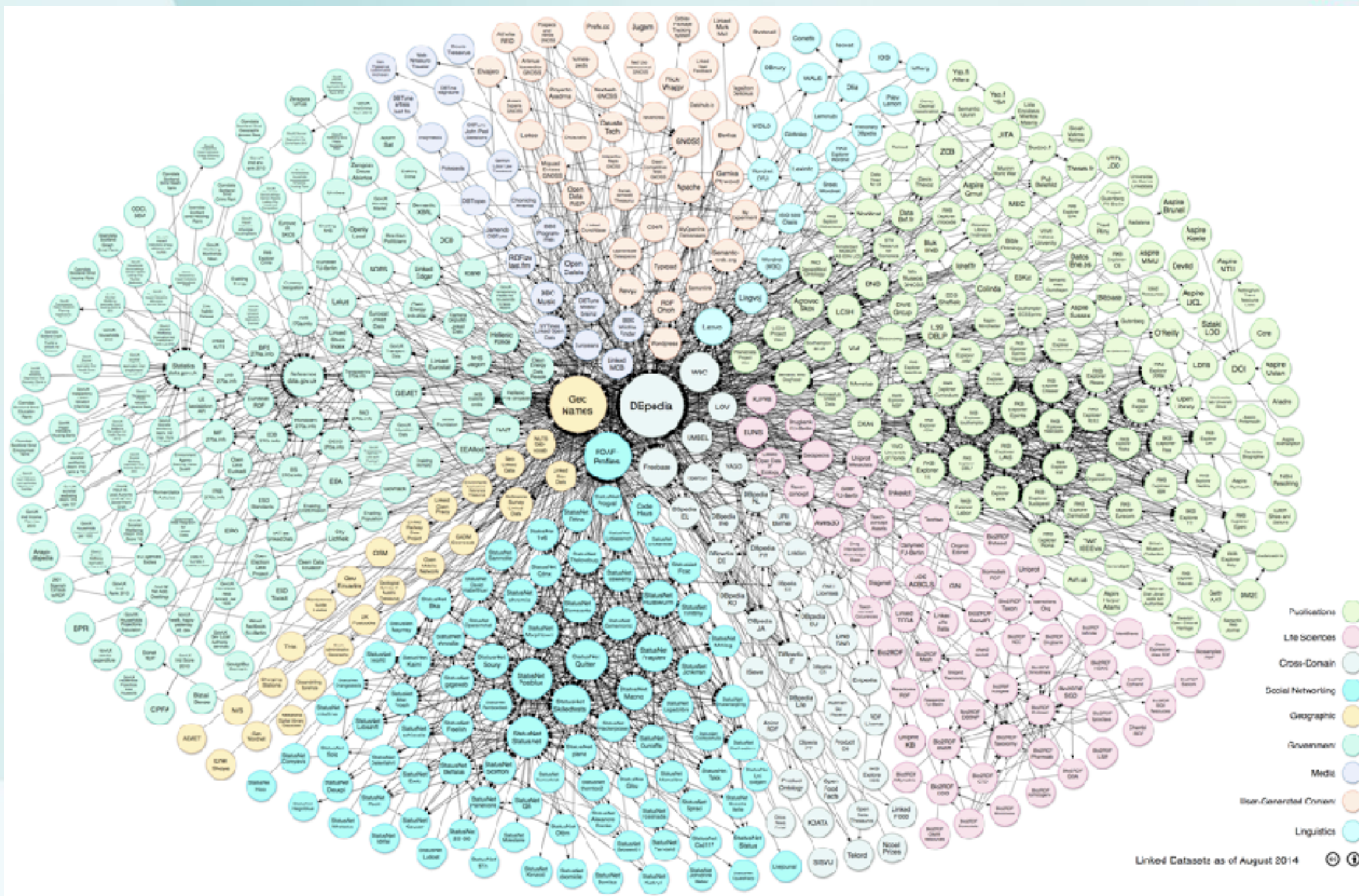
SEE HOW DATA CAN CHANGE THE WORLD

A world where knowledge creates power for the many, not the few.

A world where data frees us — to make informed choices about how we live, what we buy and who gets our vote.

A world where information and insights are accessible — and apparent — to everyone.

This is the world we choose.





➤ 4: Description Set Profile Interoperability

- Shared formal vocabularies and constraints in records

➤ 3: Description Set syntactic interoperability

- Shared formal vocabularies in exchangeable records

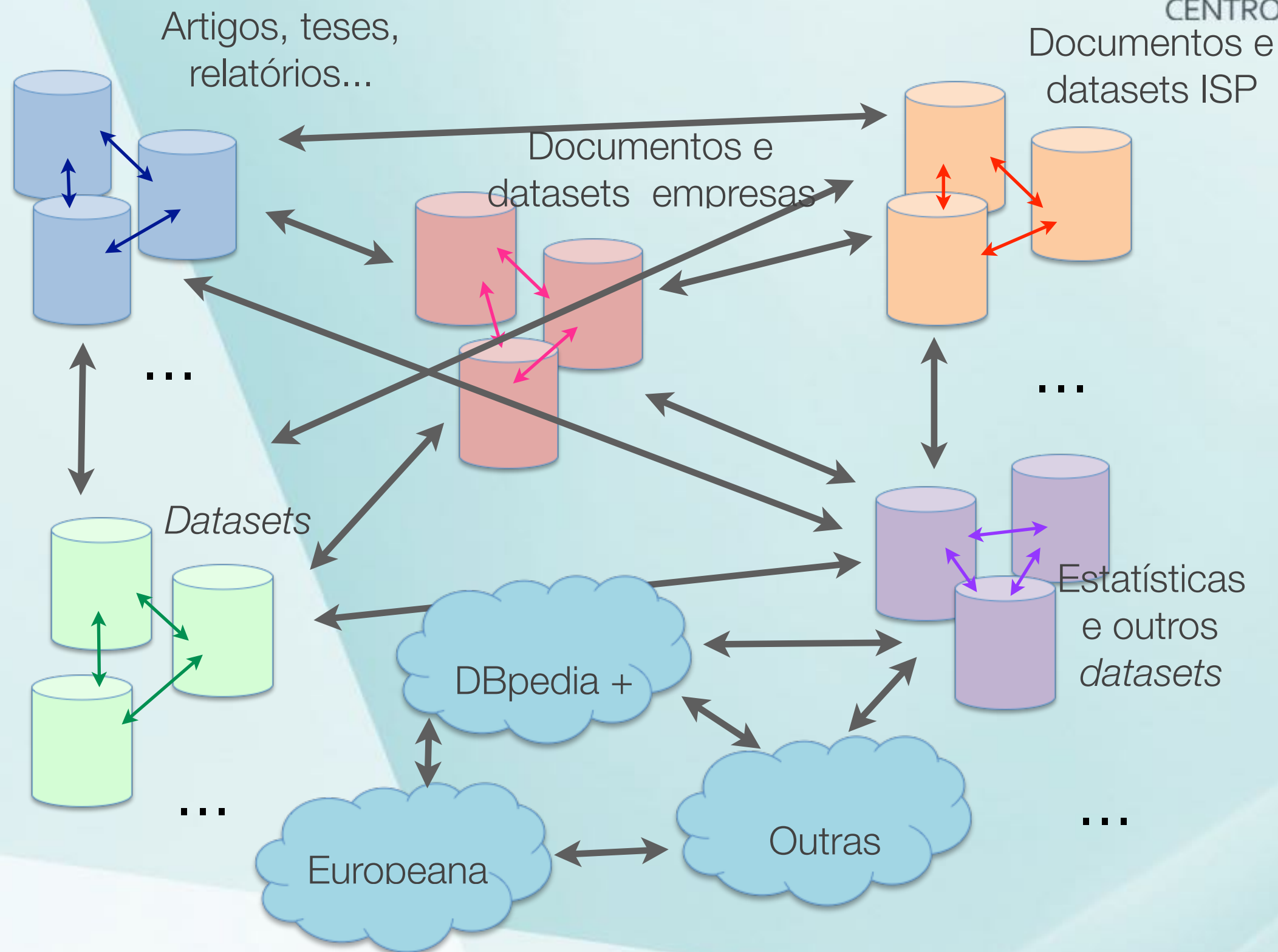
➤ 2: Formal semantic interoperability

- Shared vocabularies based on formal semantics

➤ 1: Shared term definitions

- Shared vocabularies defined in natural language

Fonte: <http://dublincore.org/documents/interoperability-levels/>





CENTROALGORITMI

mas...



CENTROALGORITMI

Rank	Place	Transport Timetables	Government Budget	Government Spending	Election Results	Company Register
24	Switzerland					
25	Italy					
26	Brazil					
26	Slovenia					
28	Korea, Republic of					

```
{
  "place": "pl",
  "dataset": "spending",
  "year": "2014",
  "score": "10",
  "rank": "16",
  "isopen": false,
  "exists": "Y",
  "digital": "Y",
  "public": "N",
  "online": "?",
  "free": "?",
  "machinereadable": "?",
  "bulk": "?",
  "openlicense": "?",
  "uptodate": "?",
  "url": null,
  "format": null,
  "licenseurl": null,
  "dateavailable": null,
  "officialtitle": null,
  "publisher": null,
  "timestamp": "2013-10-02T06:26:15.745Z",
  "submitters": [
    "Gosia Klirowicz",
    "Pierre Chrzanowski"
  ]
}
```

	Real-Time Transit	Annual Budget	Expenditure (detailed)	Election results	Air Quality	Transport Timetables	Public Facilities	Crime statistics	Procurement contracts	Food safety inspections
1 Curitiba										
2 São Luís										
3 Brasília										
4 Aracaju										
5 Belém										

	Food safety inspections	Traffic accidents	Participative Budget	Service Requests	Annual Budget	Public Facilities	Building permits	Crime statistics	Business Permits
1 Braga									
2 Guimarães									
3 Barcelos									
4 Vila Nova de Famalicão									
5 Vila Nova de Gaia									

```
<dc:type>
  workingPaper
</dc:type>
```

```
<dc:identifier>
  http://hdl.handle.net/1822/327
</dc:identifier>
```

```
<dc:language>
  eng
</dc:language>
```

```
<dc:relation>
  Documentos de Trabalho = Working Papers / NEGE ; 4/2003
</dc:relation>
```

```
<dc:relation>
  Working Papers / NEGE ; 4/2003
</dc:relation>
```

```
<dc:rights>
  open access
</dc:rights>
```

```
<dc:relation>
  http://www.sciencedirect.com/science/journal/01429612
</dc:relation>
```

```
<dc:rights>
  restrict access
</dc:rights>
```

```
<dc:publisher>
  Elsevier Science
</dc:publisher>
```

```
<dc:subject>
  Fontes de informação
</dc:subject>

<dc:subject>
  Recursos de informação
</dc:subject>

<dc:subject>
  Conceitos de informação
</dc:subject>
```

IBICT

```
<dc:type>
  article
</dc:type>
```

```
<dc:identifier>
  Pesquisa Brasileira em Ciência da Informação e Biblioteconomia, v.1,n.1, 2006.
</dc:identifier>
```

```
<dc:identifier>
  1981-0695
</dc:identifier>
```

```
<dc:identifier>
  http://hdl.handle.net/123456789/10
</dc:identifier>
```

```
<dc:language>
  por
</dc:language>
```

```
<dc:rights>
  open access
</dc:rights>
```

Cambridge



CENTROALGORITMI

```
<dc:subject>
fieldwork
</dc:subject>

<dc:subject>
Gurung
</dc:subject>

<dc:subject>
Nepal
</dc:subject>

<dc:subject>
reflections
</dc:subject>

<dc:subject>
Himalaya
</dc:subject>

<dc:subject>
change
</dc:subject>
```

```
<dc:type>
Working Paper
</dc:type>
```

```
<dc:date>
2004-05-10T12:48:42Z
</dc:date>

<dc:date>
2004-05-10T12:48:42Z
</dc:date>

<dc:date>
2002-12-19 15:45:00
</dc:date>

<dc:date>
2004-05-10T12:48:42Z
</dc:date>
```

OASIS (POSTECH)

```
<dc:rights>
BY_NC_ND
</dc:rights>

<dc:rights>
http://creativecommons.org/licenses/by-nc-nd/2.0/kr
</dc:rights>

<dc:publisher>
포항공과대학교
</dc:publisher>
```

Australia National University

```
<dc:type>
Journal article
</dc:type>

<dc:type>
Published Version
</dc:type>

<dc:identifier>
Optics Express 15.9 (2007): 5730-5741
</dc:identifier>

<dc:identifier>
1094-4087
</dc:identifier>

<dc:identifier>
http://hdl.handle.net/10440/322
</dc:identifier>
```

```
<dc:date>
2009-05-26T04:05:37Z
</dc:date>

<dc:date>
2010-12-20T06:02:26Z
</dc:date>

<dc:date>
2009-05-26T04:05:37Z
</dc:date>

<dc:date>
2010-12-20T06:02:26Z
</dc:date>
```

```
<dc:contributor>
E5122
</dc:contributor>

<dc:contributor>
u9115357
</dc:contributor>

<dc:contributor>
E2888
</dc:contributor>

<dc:contributor>
E30351
</dc:contributor>
```




CENTROALGORITMI

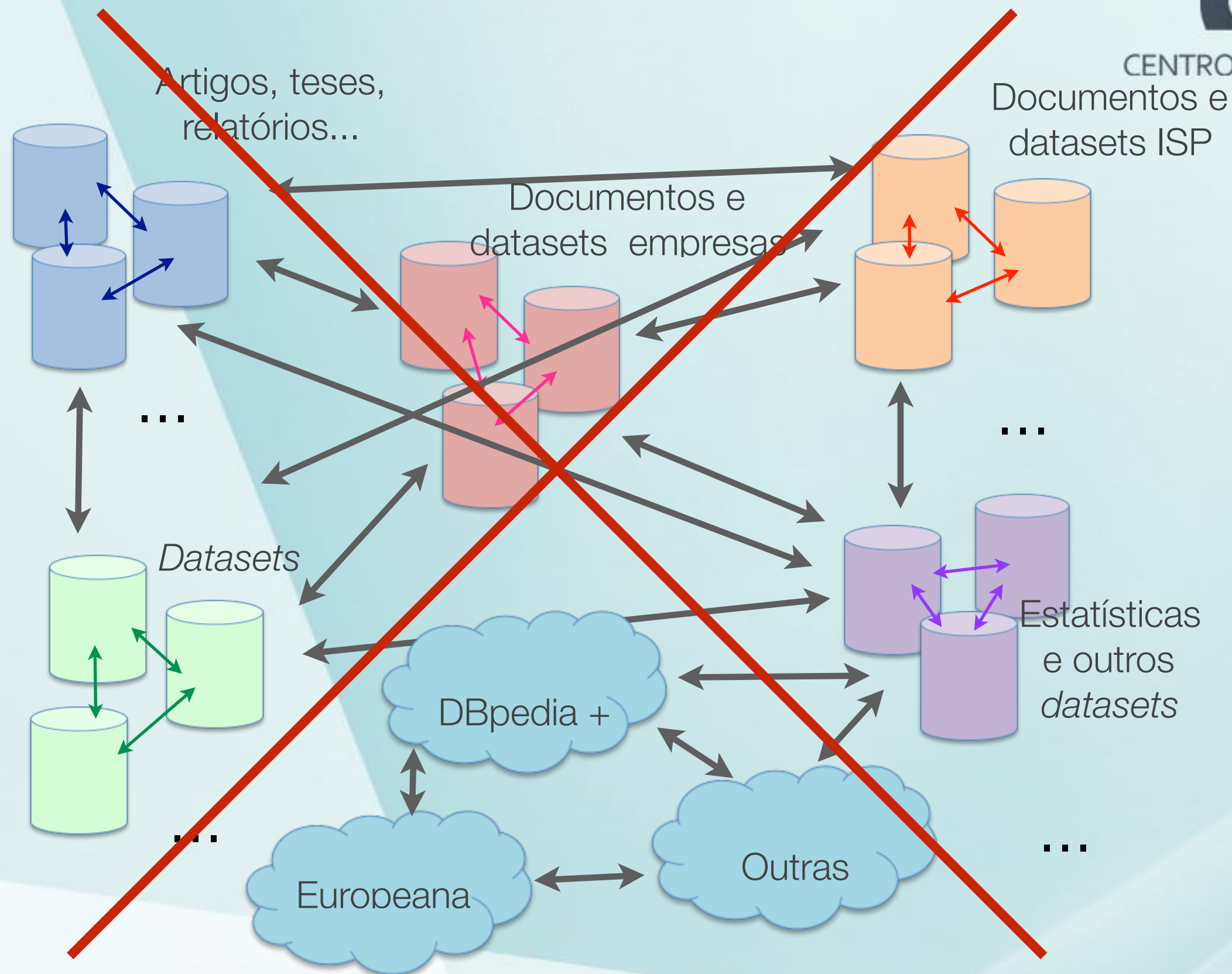
diagnóstico

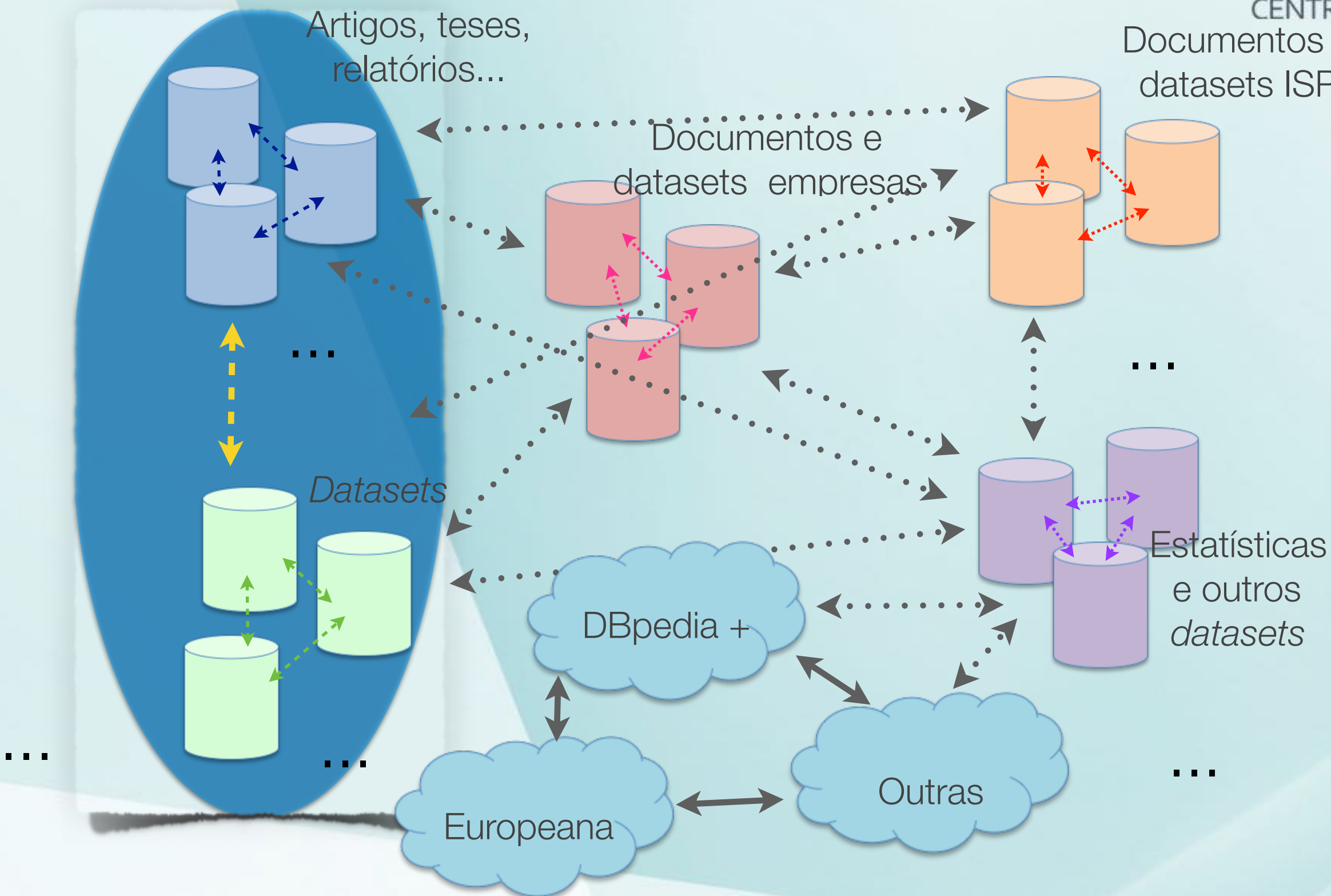
Questões

- Os **15 elementos** DC chegam para descrevermos todos os tipos de recursos sob o ponto de vista do utilizador?
- Há coerência entre os valores?
- Os valores interpretam-se todos da mesma maneira?
- Os RD interoperam entre si e com aplicações LOD?



CENTROALGORITMI





- Repositórios em **distintos níveis** de maturidade;
- Número **crescente** de repositórios;
- **Vários tipos** de recursos nos RD;
- **As mesmas propriedades** (DC simples + ...) para os descrever;
- Baixa **coerência** das descrições (vocabulários, esquemas de sintaxe).

- Mesmo os repositórios mais antigos:
 - têm problemas de interoperabilidade entre eles;
 - descrevem diferentes tipos de recursos usando as mesmas propriedades;
 - não estão preparados para interoperar na LOD cloud.

Proposta

- Diferentes propriedades para diferentes tipos de recursos;
- Utilização de esquemas de sintaxe;
- Criação, codificação em SKOS e utilização de esquemas de vocabulário;
- Preparação dos registos de metadados para integrarem a LOD cloud (linked data).



CENTROALGORITMI

ou seja



- Definição e aplicação de normas para a qualidade dos metadados;
- Criação e implementação de perfis de aplicação específicos por tipo de recurso (e eventualmente por área científica).



CENTROALGORITMI

exemplos



CENTROALGORITMI

Exemplo de guia de boas práticas - W3C

Data on the Web Best Practices

W3C Candidate Recommendation 30 August 2016



This version:

<https://www.w3.org/TR/2016/CR-dwbp-20160830/>

Latest published version:

<https://www.w3.org/TR/dwbp/>

Latest editor's draft:

<http://w3c.github.io/dwbp/bp.html>

Exemplo de vocabulário controlado em SKOS

GACSDemo

Vocabularies About Feedback Help | en Français en español in Italiano auf Deutsch

GACS Core Beta 3.1

Content language English Search

Alphabetical

Hierarchy

Groups

A B C D E F G H I J K L M N O
P Q R S T U V W X Y Z 0-9

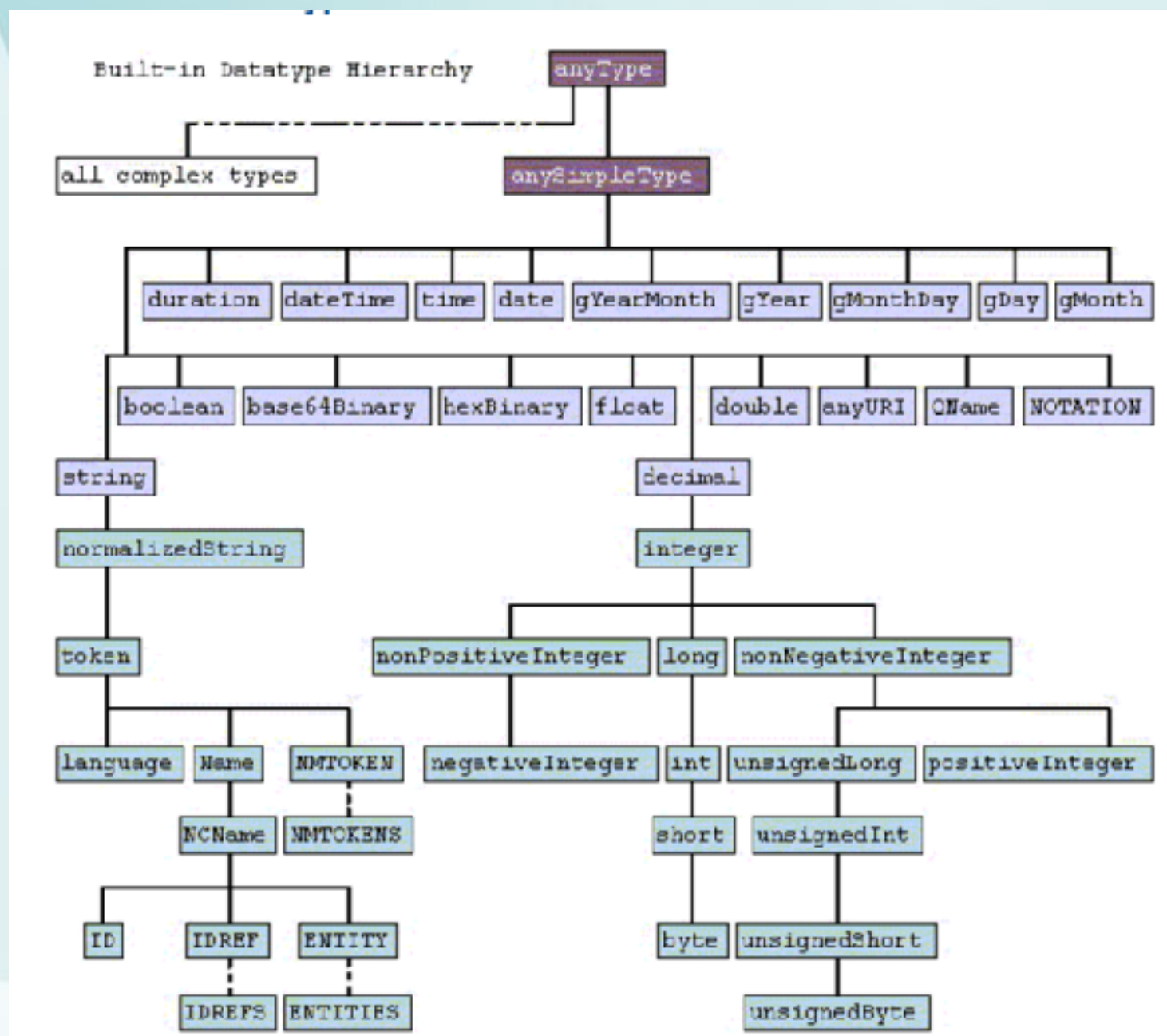
A horizons

A-horizons → A horizons
α-La (alpha-lactalbumin) → lactalbumin
ABA → abscisic acid
ABA (abscisic acid) → abscisic acid
ABA (plant hormone) → abscisic acid
Abaca → *Musa textilis*
Abaca mosaic potyvirus → Sugarcane mosaic virus
Abaca mosaic virus → Sugarcane mosaic virus
abalone → abalones
abalone mushrooms → mushrooms
abalones
Abamaceae → Liliaceae
abamectin
abandoned land
abattoir byproducts
abattoir wastes → slaughterhouse wastes
abattoirs
ABC transporters
ABC-type transporters → ABC transporters
Abcrana → Phyllophaga
abdomen
abdominal fat
Abelmoschus

Vocabulary information

TITLE	GACS Core Beta 3.1 Global Agricultural Concept Scheme Core
DESCRIPTION	The Global Agricultural Concept Scheme (GACS) is a hub for concepts related to agriculture, in multiple languages, for use in Linked Data. The idea for GACS emerged out of discussions at the World Congress of IAALD, the International Association of Agricultural Information Specialists, in July 2013. The Food and Agricultural Organization of the United Nations (FAO), CAB International (CABI), and the National Agricultural Library of the USA (NAL) agreed in October 2013 to explore the feasibility of developing a shared concept scheme by integrating their three thesauri: the AGROVOC Concept Scheme, the CAB Thesaurus (CABT), and NAL Thesaurus (NALT). In the GACS vision, the integration of these three thesauri is but the first step towards the realization of a hub that links to and from the concept schemes beyond the initial three, and in multiple language areas.
CREATOR	Global Agricultural Concept Scheme working group
CONTRIBUTOR	CAB International

Exemplo de esquemas de sintaxe (datatypes)





Exemplo Perfil de Aplicação

(isto não é uma matriz de restrições)

Elemento XML	Propriedade	Descrição	Namespace	Aplicada no âmbito de	Valores permitidos / codificação (contra-domínio)	Tipo
modified	modified	Data de modificação do esquema conceptual. Se houver mais de uma modificação, repetem-se os elementos.	http://purl.org/dc/terms/	ConceptScheme	Formato "aaaa-mm-dd" / W3CDTF/ ISO 8601	Opcional e repetível.
lifeCycleVersion	lifeCycleVersion	Versão.	http://data.opendiscoveryspace.eu/lom_ontology_ods.owl#	ConceptScheme	Texto.	Obrigatório. Apenas uma versão – elemento não repetível.
isVersionOf	isVersionOf	Link para a versão anterior.	http://purl.org/dc/terms/	ConceptScheme	IRI ¹⁰ .	Opcional. Apenas uma versão anterior – elemento não repetível.
hasTopConcept	hasTopConcept	Conceito de topo do esquema conceptual	http://www.w3.org/2004/02/skos/core#	ConceptScheme	IRI	Obrigatório e repetível.
prefLabel	prefLabel	Etiqueta preferida do conceito. Uma por idioma. Se houver mais de um idioma, repete-se o elemento. Utilização do atributo xml:lang para a identificação do idioma.	http://www.w3.org/2004/02/skos/core#	ConceptScheme	Texto. Identificação do idioma deve seguir a norma ISO 639-2 ou a ISO 639-3.	Obrigatório e repetível, mas máximo de uma prefLabel por idioma – elemento repetível apenas para prefLabels em idiomas diferentes.
altLabel	altLabel	Etiqueta alternativa sobre o conceito. Se houver mais de uma deve repetir-se o elemento. O idioma deve ser identificado.	http://www.w3.org/2004/02/skos/core#	Concept	Texto. Identificação do idioma deve seguir a norma ISO 639-2 ou a ISO 639-3	Opcional e repetível.

Exemplo Perfil de Aplicação

(codificação em XML)

```
<?xml version="1.0" ?>
<DescriptionSetTemplate xmlns="http://dublincore.org/xml/dc-dsp/2008/03/31">

  <DescriptionTemplate ID="SSEInitiative" minOccurs="1" maxOccurs="infinity" standalone="yes">
    <ResourceClass>http://purl.org/essglobal/vocab/v1.1/SSEInitiative</ResourceClass>

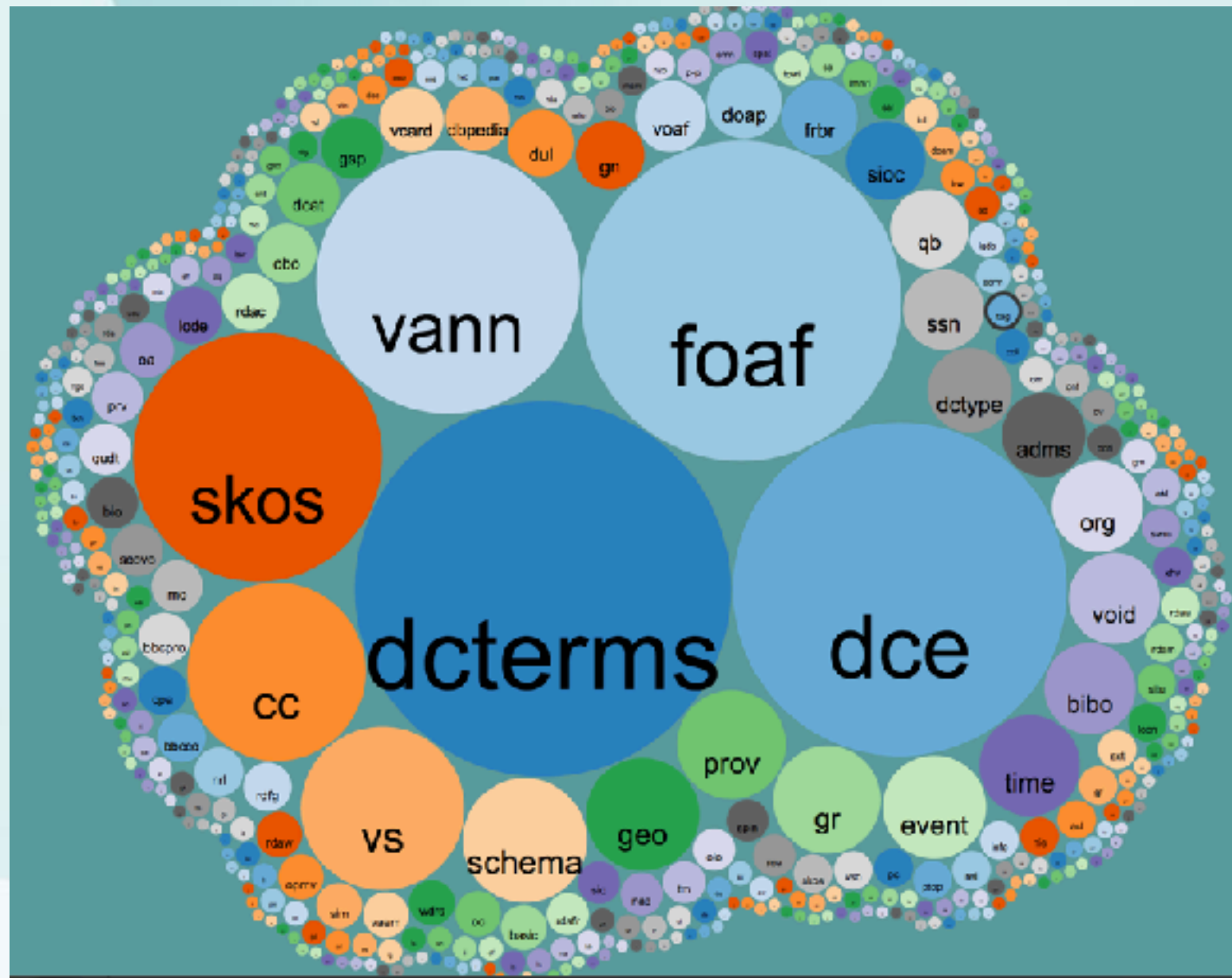
    <StatementTemplate ID="Name" minOccurs="1" maxOccurs="1" type="literal">
      <Property>http://purl.org/goodrelations/v1#name</Property>
    </StatementTemplate>

    <StatementTemplate ID="Description" minOccurs="0" maxOccurs="1" type="literal">
      <Property>http://www.purl.org/dc/terms/description</Property>
      <LiteralConstraint>
        <LanguageOccurrence>optional</LanguageOccurrence>
      </LiteralConstraint>
    </StatementTemplate>

    <StatementTemplate ID="Objective" minOccurs="0" maxOccurs="1" type="literal">
      <Property>http://purl.org/essglobal/vocab/v1.1/objective</Property>
      <LiteralConstraint>
        <LanguageOccurrence>optional</LanguageOccurrence>
      </LiteralConstraint>
    </StatementTemplate>

  </DescriptionTemplate>

</DescriptionSetTemplate>
```



Este é o tamanho da DCMI

- DCTerms -
reutilizado por 475
vocab.
- DC (simples) -
reutilizado por 375
vocab.
- DC Type -
reutilizado por 24
vocab.

Fonte: <https://lov.okfn.org>



METADATA
Dublin Core® Metadata Initiative
INNOVATION

O DC é
mencionado em 69
standards ISO

Usado no âmbito
do OAI-PMH

Standard ISO -
15836:2009

IETF RFC 5013

DC Terms em
processo de
normalização
ISO

ANSI/NISO
Standard
Z39.85-2007

Envolvimento constante em projetos das
comunidades



METADATA

Dublin Core® Metadata Initiative

INNOVATION



DCMI's work is supported, promoted and improved by Member organizations around the world:



**Metadata and Ubiquitous Access to Culture,
Science and Digital Humanities • DC**
1-4 Sep 2015 • São Paulo, Brazil

**Webinars - agora em
EN, ES e CN**

- Repositórios Digitais - área de aplicação extremamente relevante para a DCMI;
- Datasets (não apenas científicos) - foco cada vez mais relevante a nível global.

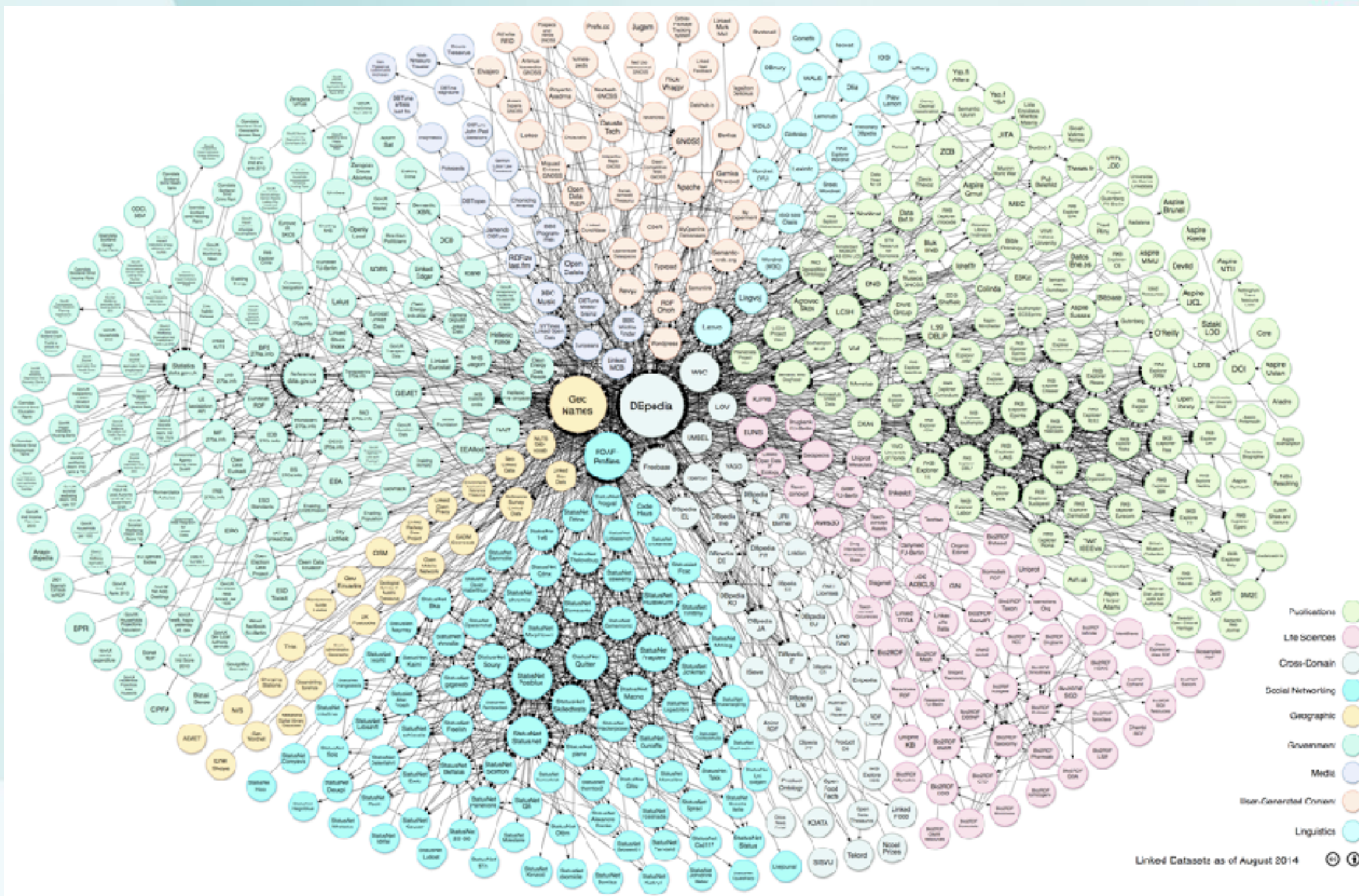
- Há áreas do globo subrepresentadas na DCMI - a América Latina é uma delas;
- O Brasil é motor na América Latina;
- Queremos o Brasil e outros países da AL envolvidos, a contribuir, na DCMI.

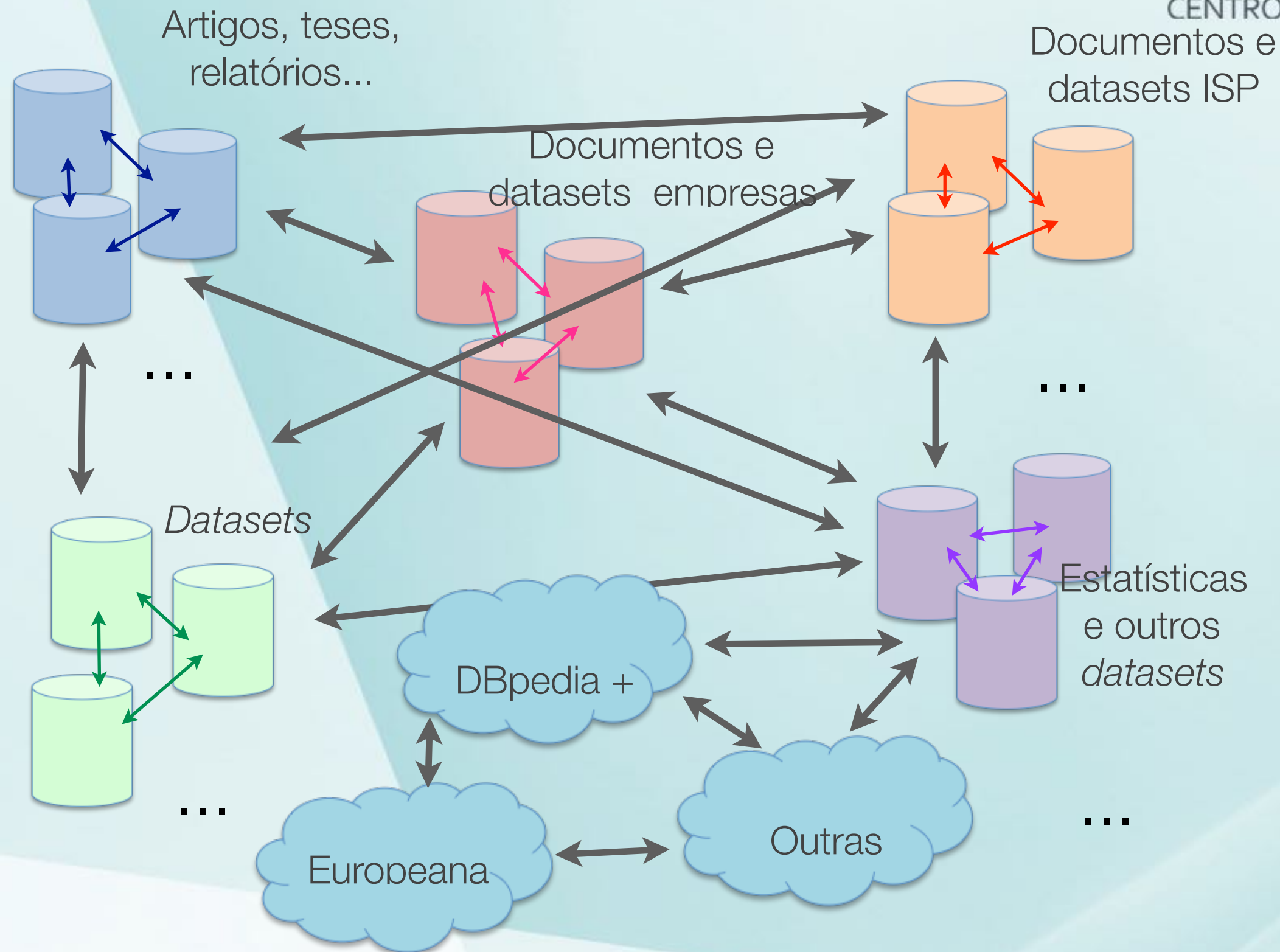
Desafios

- Criar grupos no Brasil que, em conjunto com pessoas de outros países desenvolvam:
- guias de boas práticas para a codificação de metadados em Repositórios Digitais;
- vocabulários controlados em SKOS;
- Perfis de Aplicação específicos.

porque...

- A DCMI **acolhe e dá visibilidade** a iniciativas de metadados para Repositórios Digitais;
- “Começar pequeno”, mas ir longe - a DCMI ajuda a **estabelecer relações** com outros grupos e pessoas.
- A DCMI é o **ponto de encontro** das comunidades de metadados.







CENTROALGORITMI



@DublinCore



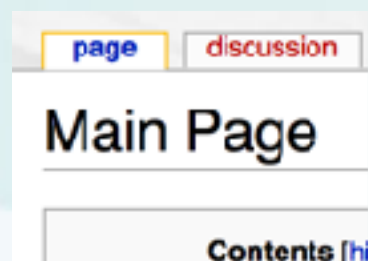
Dublin Core - Ibero-América



DC-General



dublincore.org



DCMI Wiki



CENTROALGORITMI

Obrigada.



FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR Portugal

Boa Mistura, Luz nas Viegas, São Paulo, Brasil